Coláiste na Tríonóide, Baile Átha Cliath
Trinity College Dublin
Ollscoil Átha Cliath | The University of Dublin

Information Theoretical Aspects of Complex Systems

Lecture 2.07

EEU45C09 / EEP55C09
Self Organising Technological Networks

CONNECT
Networks of the Future

Nicola Marchetti
nicola.marchetti@tcd.ie

Measuring complexity

- What characterises a complex symbol sequence / pattern ?
- There are many suggestions on how one should quantify complexity
- What quantity to use depends on what we are looking for in the system under study
- We will focus on quantities related to *how correlation information is distributed* in the system
- Approach suggested by Peter Grassberger in the 1980's
- Let us consider two of the examples discussed last time, the completely ordered ("crystal") sequence and the completely random ("gas") sequence:

...0101010101010101010101010101...

...1100001101000101100100110111 01...

Simple vs Complex

- In the first case, the entropy is minimal (0 bits), and in the second case is at its maximum (1 bit)

- None of these are typically considered as complex, even though in some contexts (e.g., Algorithmic Information Theory) the random one has been called complex

- The precise configuration of the gas sequence needs a lot of information to be specified, but the system that generates it is simple: “toss a coin for ever…”

- The crystal is also generated by a system which has a simple description: “repeat 01…”

Simple vs Complex

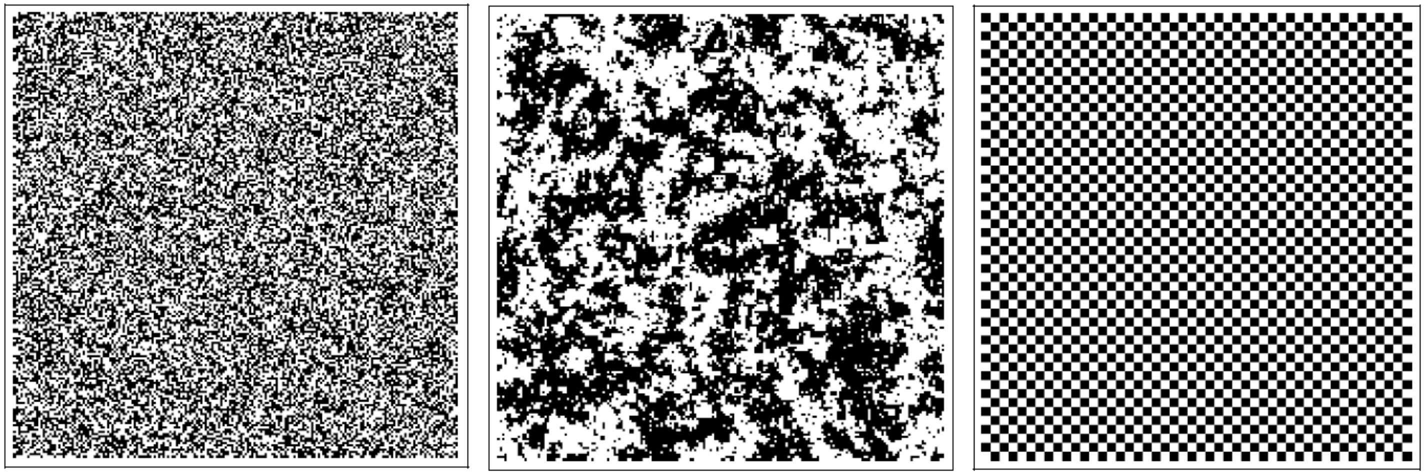


Fig. 1

- We are usually not considering a fully random pattern (left) as complex. Neither is a fully ordered structure exemplified by the checkerboard pattern (right) considered as complex

- The potential to exhibit complex characteristics usually requires patterns in between these extremes, patterns that may contain some order in correlations, but where *correlations may reach over larger distances*, and where there is also some randomness as well

- Here this is illustrated by a typical state in a physical spin system close to the critical temperature (middle)

Complexity as long-range correlation

- If neither the most ordered nor the most random systems are considered complex, we should look for a measure of complexity that can take high values when the entropy is somewhere in between, see Fig. 1

- One information-theoretic characteristic that has been considered as an important component for a complex system is to what extent correlation information is spread out in the system

- If there is a lot of information in long-range correlation, it may be more difficult to analyse and describe the system

Correlation complexity

- Grassberger suggested a quantity, which can be defined as a weighted sum of information contributions from different block lengths

$$\eta = \sum_{m=1}^{\infty} (m-1)\, k_m \qquad \textbf{(1)}$$

- We call this the *correlation complexity*, and it can be rewritten (if $k_{corr} > 0$) as

$$\eta = k_{\text{corr}} \sum_{m=1}^{\infty} (m-1) \frac{k_m}{k_{\text{corr}}} = k_{\text{corr}}\, \overline{(m-1)} = k_{\text{corr}}\, d_{\text{corr}}$$

Correlation complexity

- We have introduced an *average correlation distance* d_{corr}

- This is based on the average block length at which correlation information is found, but where correlation distance is defined to be one less than the block length (or the "distance" between first and last symbol of the block)

- This complexity measure is non-negative, but it is not unbounded. For the "crystal" example $\eta = 1$ and for the totally random sequence $\eta = 0$

- It also turns out that the correlation complexity can be interpreted as the average information contained in a semi-infinite symbol sequence $(\ldots x_{-2}, x_{-1}, x_0)$ about its continuation $(x_1, x_2, x_3, \ldots)$

Correlation complexity

- This quantity can be written as a Kullback information

- Suppose first that the preceding sequence, that we may observe, has a finite length m and we denote it as

$$\sigma_m = (x_{-m+1}, \ldots, x_0)$$

- The continuation has finite length n and we denote it as

$$\tau_n = (x_1, \ldots, x_n)$$

- At the end we shall look at infinite limits of m and n

Correlation complexity

- The a priori description of the continuation is given by probabilities p(τ_n), but after we have observed the preceding sequence σ_m , we can replace that with the conditional probabilities p($\tau_n|\sigma_m$)

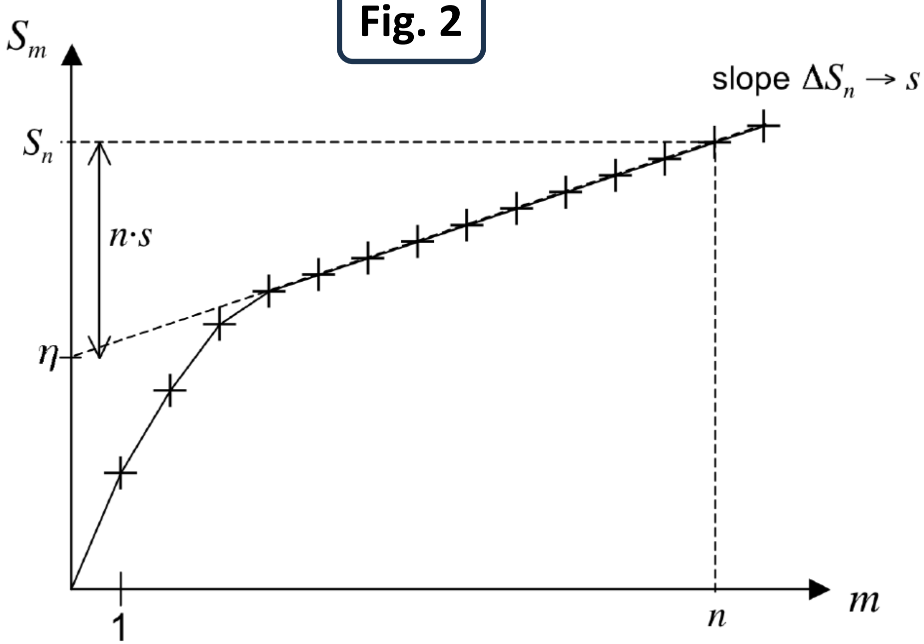
- The information we gain by this is a Kullback information, and we take the average over all possible preceding sequences together with the limit of infinite lengths

$$\eta = \lim_{m\to\infty}\lim_{n\to\infty}\sum_{\sigma_m} p(\sigma_m)\sum_{\tau_n} p(\tau_n \mid \sigma_m)\log\frac{p(\tau_n \mid \sigma_m)}{p(\tau_n)} \geq 0 \qquad \textbf{(2)}$$

Correlation complexity

❑ We can also relate the correlation complexity to the block entropies, see Fig. 2 (we already saw this figure in Lecture 2.05)

Fig. 2



The block entropy S_m tends to an asymptotic line with slope equal to the entropy s

In the limit of infinite sequences, $S_m - m \cdot s$ is the correlation complexity η, and graphically, it is the intersection point on the vertical axis of the asymptotic line for the block entropy

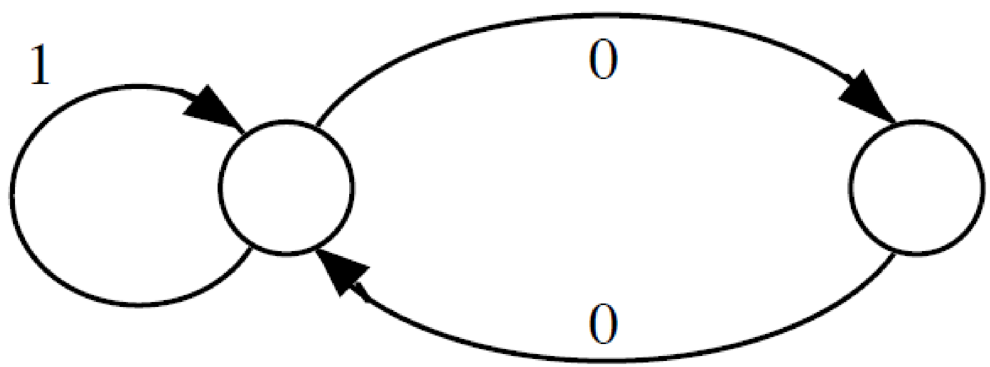
$$\lim_{m\to\infty}(\eta + \mathrm{m}\cdot \mathrm{s}) = S_m \quad \Rightarrow \quad \boxed{\eta = \lim_{m\to\infty}(S_m - m\,s)}$$

Correlation complexity for Markov processes/models

- For a *Markov process* the correlation complexity η is easy to calculate since there is no correlation information from blocks longer than 2, and therefore we get that $\eta = k_2$ (see Eq. (1))

- For a *hidden Markov model*, the situation is different

- As we saw in Lecture 2.06, the system below may have correlation information in arbitrarily long blocks of symbols



Correlation complexity for Markov processes/models

- If almost all past sequences σ_m, in the infinite limit, determine whether we are in the left (L) or in the right (R) node, we can rewrite Eq. (2)

- The probability of the future sequence τ_n is given by the node we are in, so we can group together all σ_m leading to the left node L (and in the same way all σ_m leading to the right one, R). Then we have

$$\eta = \lim_{n \to \infty} \sum_{z \in \{L,R\}} p(z) \sum_{\tau_n} p(\tau_n \mid z) \log \frac{p(\tau_n \mid z)}{p(\tau_n)} \qquad \text{(3)}$$

- We now use the definition of conditional probability $p(\tau_n|z)$ to rewrite the argument in the logarithm

$$\frac{p(\tau_n \mid z)}{p(\tau_n)} = \frac{p(\tau_n , z)}{p(\tau_n)\, p(z)} = \frac{p(z \mid \tau_n)}{p(z)}$$

Correlation complexity for Markov processes/models

- We can then rewrite Eq. (3) as follows

$$\eta = \lim_{n\to\infty} \sum_{z\in\{L,R\}} p(z) \sum_{\tau_n} p(\tau_n \mid z) \log \frac{p(z \mid \tau_n)}{p(z)} =$$

$$= \lim_{n\to\infty} \sum_{z\in\{L,R\}} p(z) \sum_{\tau_n} p(\tau_n \mid z) \left(\log \frac{1}{p(z)} + \log p(z \mid \tau_n) \right) =$$

$$= \sum_{z\in\{L,R\}} p(z) \log \frac{1}{p(z)} - \lim_{n\to\infty} \sum_{\tau_n} p(\tau_n) \sum_{z\in\{L,R\}} p(z \mid \tau_n) \log \frac{1}{p(z \mid \tau_n)}$$

Prove this.

- The last sum over nodes z is the entropy of which node we were in, conditioned on observing the future sequence τ_n
- But the starting node for the future sequence is almost always uniquely given by the sequence τ_n
- Only when there are only zeroes in τ_n, we do not know, but that happens with probability 0 in the infinite limit
- Therefore, there is no uncertainty of z given the future τ_n, and the last entropy term is 0

Correlation complexity for Markov processes/models

- Finally, we get

$$\eta = \sum_{z \in \{L,R\}} p(z) \log \frac{1}{p(z)}$$

- And we can conclude that, for this situation, the correlation complexity equals the entropy of the stationary distribution of the states in the finite state automaton describing the hidden Markov model

- Note, though, that this does not hold in general, but only for certain types of hidden Markov models

Extension to higher dimensions

- The decomposition of information into entropy and contributions from different correlation lengths can be extended to lattice systems of any dimension

- We will now focus on the 2D case, but formalism can be extended to higher dimensions

- Consider an infinite 2D lattice in which each site is occupied by a symbol from a finite alphabet (e.g., 0 or 1)

- We can calculate (or at least estimate) the relative frequencies, with which finite configurations of symbols occur in the lattice

- Let $A_{M\times N}$ be a specific ($M\times N$)-block occurring with probability $p(A_{M\times N})$

Extension to higher dimensions

- Then the entropy, i.e., the average information per site, is

$$s = \lim_{M,N\to\infty} \frac{1}{MN} S_{M\times N}$$

- Where the block entropy $S_{M\times N}$ is defined by

$$S_{M\times N} = \sum_{A_{M\times N}} p(A_{M\times N}) \log \frac{1}{p(A_{M\times N})}$$

Extension to higher dimensions

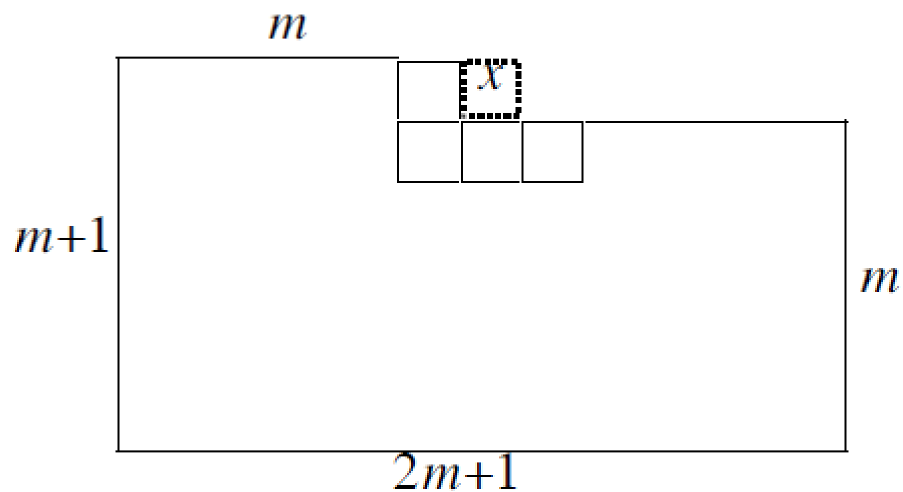


Fig. 3

- Let B_m be a certain configuration of symbols arranged as follows: m rows of symbols, each of length 2m+1, are put on top of each other, and on the m first symbols of the top row a sequence of m symbols is placed, see Fig. 3

- We also introduce the notation $B_m x$ for the configuration that adds the symbol x to B_m after the m-th symbol in the top row

Extension to higher dimensions

- Then we can introduce the conditional probability for a certain character x , given that we have already observed the characters in the configuration B_m

$$p(x \mid B_m) = \frac{p(B_m x)}{p(B_m)}$$

- This can be interpreted as the conditional probability for the "next" character given that we have seen the "previous" m+m(2m+1) = 2m(m+1) characters

- The average entropy is

$$H_m = \sum_{B_m} p(B_m) \sum_x p(x \mid B_m) \log \frac{1}{p(x \mid B_m)}$$

Extension to higher dimensions

- For m = 0, we define $H_0 = S_{1\times 1}$, or the entropy of the single character distribution

- One can prove, that in the limit $m \to \infty$, that H_m is equal to the entropy (as it was in the 1D case)

$$s = \lim_{m \to \infty} H_m = H_\infty$$

- As in the 1D case, the average information of $\log_2(2)$ = 1 bit, per lattice site, can be decomposed into a term quantifying the information in correlations from different lengths (including density information) and a term quantifying the internal randomness of the system

$$1 = k_{\mathrm{corr}} + s$$

Extension to higher dimensions

- As the density information k_1 does not depend on the dimensionality, it will be expressed as in the 1D case

$$k_1 = \sum_x p(x) \log \frac{p(x)}{1/2} = 1 - S_{1\times1} = 1 - H_0$$

- Extending this reasoning to length m , we can define the correlation information over length m by the difference between two consecutive estimates of the entropy s , i.e., $k_{m+1} = -H_m + H_{m-1}$

- Let us now introduce an operator R that reduces a configuration B_m to a configuration $B_{m-1} = RB_m$, by taking away the symbols from the leftmost and rightmost columns as well as from the bottom row

Extension to higher dimensions

- Then k_m can be written as the average Kullback information when the distribution for "next" character, given a conditional configuration B_m , replaces an a priori distribution with a smaller conditional configuration $B_{m-1} = RB_m$,

$$k_{m+1} = -H_m + H_{m-1} = \sum_{B_m} p(B_m) \sum_x p(x \mid B_m) \log \frac{p(x \mid B_m)}{p(x \mid \mathrm{R}B_m)} \geq 0$$

- The correlation information k_{corr} is then decomposed by

$$k_{\mathrm{corr}} = \sum_{m=1}^{\infty} k_m$$

Extension to higher dimensions

- This procedure can be repeated in higher dimensions

- Note: the definition of correlation information in 2D contains a choice of direction and rotation: the B_m configuration can be chosen in 8 different ways – **Why?**

- In 1D there are only two ways to choose, direction left or right, but the choice does not change the definitions

- Furthermore, and this holds for both 1D and higher dimensions, one may use other ways to decompose total redundancy which result in different types of correlation measures

Ackowledgement

- Kristian Lindgren, "Information Theory for Complex Systems", pages 30-36